BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 6: e90989
doi: 10.3897/biss.6.90989

OPEN ACCESS

Conference Abstract

Early Career Scientists are Critical to the FAIR Data Pathway

Sara E Hansen‡, Nico Franz§, Anna K Monfils‡

‡ Central Michigan University, Mount Pleasant, United States of America
§ Arizona State University, Tempe, United States of America

Corresponding author: Sara E Hansen (hanse2s@cmich.edu)

Received: 28 Jul 2022 | Published: 01 Aug 2022

Citation: Hansen SE, Franz N, Monfils AK (2022) Early Career Scientists are Critical to the FAIR Data Pathway. Biodiversity Information Science and Standards 6: e90989. https://doi.org/10.3897/biss.6.90989

Abstract

Biodiversity data that meet the principles of Findability, Accessibility, Interoperability, and Reusability (FAIR) have tremendous potential to facilitate integrative science addressing large-scale biodiversity challenges. In order to increase the amount, breadth, and quality of FAIR biodiversity data, we must look to the individuals responsible for creating, sharing, and publishing their data. An open conversation and collaboration among those generating data and those using, extending, and curating data is critical (Leonelli 2016). Graduate students and other early career scientists often work between these points along the data pathway. As they enter the field of biodiversity science, they are required to have not only knowledge of taxa, ecosystems, and research practices but also a new set of skills to handle data of many types (Monfils et al. 2021). As best practices in biodiversity data science are actively defined and the need for FAIR data and Open Science continues to increase, involvement by early career scientists is most valuable, as they share their cutting-edge skills to integrate and publish biodiversity data. Early career scientists are often more likely and willing to publish the data they collect (Campbell et al. 2019) and can directly benefit through peer-reviewed publications and professional connections. However, safeguards and incentives are necessary, as data publication can carry a heavy burden and pose threats to emerging careers. Early career scientists may spend large amounts of time and effort to publish data without recognition, expose potential faults in their methodologies that would not otherwise be apparent, lose ownership of their data and derivatives, or have their data used unethically or inappropriately (Franz and Sterner 2018

© Hansen S et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

). Despite a demonstrated positive attitude among early career scientists toward Open Science and Open Data practices and values (Toribio-Flórez et al. 2021), the responsibility and risks of data sharing cannot be overlooked.

In working to improve and expand the FAIR data landscape and support future biodiversity data science careers, there are several potential challenges to consider. First, there are often barriers to acquiring the knowledge and skills needed to utilize, aggregate, and publish data. Teaching of biodiversity data science has not been widely adopted in undergraduate curricula (Ellwood et al. 2019, Emery et al. 2021), so scientists must seek out educational materials on their own during their graduate or postgraduate careers. Integrating biodiversity data science into all stages of science careers is necessary; for example, by introducing incoming college students to the full range of roles in biodiversity science that include data, incorporating data skills early in undergraduate programs and emphasizing the importance and value of biodiversity data alongside established life science concepts, increasing opportunities to gain data expertise in graduate programs regardless of where students' career interest lies along the data continuum, and supporting established professionals who recognize the need for these skills and efforts without their own formal preparation and institutional resources (Barone et al. 2017). Training materials created for biodiversity data science professionals, including those offered by the Carpentries, Global Biodiversity Information Facility (GBIF), Biodiversity Information Standards (TDWG), and others can be translated to all these levels. Second, sharing FAIR data is not worth the effort or risk for many individuals. If graduate programs and future employers do not place a high value on biodiversity data science and do not train, fund, and promote early career scientists in this area, there is little incentive to pursue this expertise at all. Because early career scientists are often encouraged to publish data without clearly defined pathways or widely available resources, they are vulnerable to mistakes and criticism that could harm their careers. Finally, conversations and initiatives in biodiversity data science may not be accessible to early career scientists; appropriate on-ramps are needed to allow both novices and experts in this field to meaningfully participate. By taking advantage of these connections and opportunities, early career scientists can promote their work and advocate for their needs, and established professionals can gain the perspective necessary to successfully guide the early career workforce toward specific goals relative to FAIR data curation and publication. Here we will present our experience in each of these areas and share the ways in which we can continue to support and advance early career scientists in biodiversity data science.

Keywords

Biodiversity data science, Open Science, Open Data

Presenting author

Sara E Hansen

Presented at

TDWG 2022

References

- Barone L, Williams J, Micklos D (2017) Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators. PLoS Computational Biology 13 (10): e1005755. https://doi.org/10.1371/journal.pcbi.1005755
- Campbell H, Micheli-Campbell M, Udyawer V (2019) Early Career Researchers Embrace Data Sharing. Trends in Ecology & Evolution 34 (2): 95-98. https://doi.org/10.1016/j.tree.2018.11.010
- Ellwood ER, Sessa JA, Abraham JK, Budden AE, Douglas N, Guralnick R, Krimmel E, Langen T, Linton D, Phillips M, Soltis PS, Studer M, White LD, Williams J, Monfils AK (2019) Biodiversity Science and the Twenty-First Century Workforce. BioScience 70 (2): 119-121. https://doi.org/10.1093/biosci/biz147
- Emery NC, Crispo E, Supp SR, Farrell KJ, Kerkhoff AJ, Bledsoe EK, O'Donnell KL, McCall AC, Aiello-Lammens ME (2021) Data Science in Undergraduate Life Science Education: A Need for Instructor Skills Training. BioScience 71 (12): 1274-1287. https://doi.org/10.1093/biosci/biab107
- Franz NM, Sterner BW (2018) To increase trust, change the social design behind aggregated biodiversity data. Database 2018 https://doi.org/10.1093/database/bax100
- Leonelli S (2016) Data-Centric Biology: A Philosophical Study. University of Chicago Press [ISBN 9780226416335] https://doi.org/10.7208/chicago/9780226416502.001.0001
- Monfils AK, Krimmel ER, Linton DL, Marsico TD, Morris AB, Ruhfel BR (2021) Collections Education: The Extended Specimen and Data Acumen. BioScience 72 (2): 177-188. https://doi.org/10.1093/biosci/biab109
- Toribio-Flórez D, Anneser L, deOliveira-Lopes FN, Pallandt M, Tunn I, Windel H (2021) Where Do Early Career Researchers Stand on Open Science Practices? A Survey Within the Max Planck Society. Frontiers in Research Metrics and Analytics 5 https://doi.org/10.3389/frma.2020.586992